Energy & Buildings 215 (2020) 109864

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data

Chengliang Xu, Huanxin Chen*

Department of Refrigeration and Cryogenic, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 29 October 2019
Revised 27 December 2019
Accepted 12 February 2020
Available online 29 February 2020

Keywords:
Deep learning
Quantile regression
Anomaly detection
Building energy management

ABSTRACT

With the development in information technologies, today's building energy consumption can be well monitored by the building energy management systems. However, in most real applications there is no clear definition of abnormal building energy consumption. To overcome this limitation, this work proposes a novel deep learning based unsupervised anomaly detection framework that includes recurrent neural networks and quantile regression. Moreover, this framework is able to produce a prediction interval to detect and evaluate abnormal building energy consumption. The framework has been applied to analyze the energy data collected from three different residential houses, and anomaly detection results are evaluated by the quantile regression range. The research results can provide promising solutions for building managers to detect abnormal energy performance, and is also valuable to assess the level of anomalies and spot opportunities in energy conservation.

1. Introduction

From 2001 to 2015, the energy consumption in the Chinese civil building sector increased with an average annual growth rate of 11.82%. While in 2015, the energy consumption has reached a total value of 857 million tons of standard coal equivalent, which account for about 20% of all the energy consumption in China [1]. However, it is estimated that 16% of the energy consumed during building operations is wasted as widespread existence of improper residential energy consumption behavior and control strategies [2]. To mitigate the energy waste during building operation, one of the most commonly used solutions is anomaly detection.

Anomaly detection, as the name implies, is the process of identifying an observation that deviates greatly from other observations in a data set. In the field of building energy consumption, anomaly is equal to abnormal energy consumption, which is generally caused by equipment faults or improper operation. In recent years, data-driven approaches are widely adopted to achieve reliable and robust anomaly detection for building energy consumption. Wang et al. [3] summarize anomalies into two levels according to the scope of research object: whole building level, system and component level. The whole building anomaly detection only focus on the overall consumption of the building, while system and component anomaly detection can identify and locate exactly which component or subsystem leads to anomalous issues. Despite the simplicity of aforementioned anomaly classification, a more common concern during anomaly detection is that whether there are explicit labels for anomalies. Therefore, we divided existing anomaly detection related articles in the building field into two categories, i.e., anomaly detection with anomaly labels and anomaly detection without anomaly labels.

1.1. Anomaly detection with anomaly labels

In general, anomaly detection with anomaly labels in building field focuses on finding anomalous behavior in system or individual equipment. System anomalies can be determined by energy consumption, and equipment anomalies can be determined by equipment related parameters. Essentially, the aim of anomaly detection with anomaly labels is to obtain high-precision classification models. Capozzoli et al. [4] used CART method and artificial neural networks and basic ensembling method to automatically detect outliers in building energy consumption, and results showed that this method can improve fault detection process by reducing the false anomalies. Araya et al. [5] proposed an ensemble learning framework for anomaly detection in building energy consumption. A pattern-based anomaly classifier is firstly applied in different test datasets, and then the hypotheses of all the anomaly classifiers are combined to create a new ensemble anomaly classifier, and it was reported that the proposed method is suitable both

* Corresponding author.
E-mail address: chenhuanxin@tsinghua.org.cn (H. Chen).

https://doi.org/10.1016/j.enbuild.2020.109864
0378-7788/

Nomenclature

ANN	artificial neural network
ARIMA	autoregressive integrated moving average
CART	classification and regression tree
BIC	Bayesian information criterion
CUSUM	cumulative sum
DT	decision tree
EWMA	exponentially weighted moving average
GESD	generalized extreme studentized deviate
HQC	Hannan and Quinn information criterion
HJC	Hatemi-J criterion
KSC	K-Spectral centroid clustering algorithm
PICP	prediction interval coverage probability
PINAW	prediction interval normalized average width
QR	Quantile regression
RNN	recurrent neural network
SD	standard deviation
SVR	support vector regression
VRF	variable refrigerant volume

for stringent anomaly detection and demanding false positive requirements.

As creating energy consumption datasets with anomaly labels on whole building level is onerous and expensive, most papers focus on the anomaly detection on system and equipment level [6–8]. Guo et al. [6] used deep belief network to diagnose four kind of faults of the variable flow refrigerant system, and the fault diagnosis correct rate of the proposed model is 97.7%, which indicates that deep belief network is suitable to intelligently diagnose VRF system faults and maintain it in a timely manner. Li et al. [7] proposed a three-stage fault diagnosis method combining DT model with virtual sensor-based fault indicators for practical VRF systems, three faults air-side fouling, refrigerant undercharge and overcharge are selected to validate the proposed method, and the fault diagnosis accuracies can reach 94.44% for the experimental datasets. For diagnosis of the energy performance of VRF systems, Liu et al. [8] adopted SVR and EWMA control chart to diagnose the system energy uses, two other methods $\bar{x}-R$ and CUSUM control charts are also implemented to verify the reliability of the proposed method.

1.2. Anomaly detection without anomaly labels

For practical applications, in contrast, anomaly detection without anomaly labels is closer to the actual requirements of building operation managers, as it is expensive and time consuming to generate labelled data for abnormal energy consumption in buildings. Therefore, a promising way to solve this dilemma is to find an algorithmic way of labeling anomalies. Chou and Telaga [9] applied a two sigma rule to detect anomaly, and this rule simply classifies any points outside of 2 SD from the mean as anomalous data. More complicated statistical methods such as Generalized extreme studentized deviate (GESD) algorithm proposed by Rosner [10], has been implemented in anomaly detection in building energy consumption and proved to be computationally efficient in handling masses of building energy data [11,12], and Fan et al. [13] applied GESD algorithm to identify anomalies for features in clusters, results show that the application of GESD leads to reduced computation load and improved prediction performance. Other statistical methods like k-means clustering algorithm [14] and principal component analysis-based methods [15] are also successfully used to perform anomaly detection in the building field.

Moreover, machine learning techniques have gained increasing attention in anomaly detection in recent years [16], and machine learning based anomaly detection methods can be further classified into unsupervised learning-based methods and regression-based methods [17]. Commonly used unsupervised learning algorithms include association rules and clustering, Sun et al. [18] used association rule mining to create fault detection thresholds for finding anomalies. Yu et al. [19] also used association rule mining based method to discover all associations and correlations between building operational data, and energy waste in the air-conditioning system as well as equipment faults can be directly identified by this method. Cabrera and Zareipour [20] applied association rule mining to extract association rules and explore lighting waste patterns in educational institute. Fan et al. [21] applied an unsupervised learning method autoencoders to detect anomalies in building energy data. Regression-based methods are mainly used to develop benchmarking models for anomaly detection. Mavromatidis et al. [22] used an artificial neural network to predict the energy benchmarking values, and defined the actual value that exceeds the predicted value by 10% or more as anomalies. Yan et al. [23] applied the back-propagation neural network algorithm to predict the performance of ground source heat pump system, and anomalies are detected by using the interquartile range rule. Zhao et al. [24] adopted SVR to develop the benchmarking performance of the heat transfer efficiency of the sub-cooling section, then the residuals between the current performance indexes and benchmark values were calculated to detect the anomalies. Liu and Nielsen [25] introduced periodic auto-regression with exogenous variables model to detect daily pattern anomalies, and Euclidean distance and KSC distance are used to define the scope of the anomaly.

To summarize, anomaly detection methods have been successfully applied in the field of detecting abnormal building energy consumption. However, for anomaly detection without anomaly labels, statistical methods and machine learning methods mentioned above have some inherent disadvantages. For statistical methods, anomaly is generally defined as point outside of several standard deviations from the mean. This kind of definition is straightforward and intuitive, but it may lead to the roughly classification of high energy consumption, and the complexity of residential electricity consumption patterns would be ignored. For machine learning methods, anomaly is generally defined as point outside the boundaries based on the predicted value, and in this context, the position of the boundaries is not fixed. However, the above methods can only judge whether the object is an outlier, while the severity of the outlier is not evaluated. Due to the above challenges, anomaly detection of building energy consumption is still the research hotspot at present.

Therefore, to fill the gaps in previous research, an energy consumption anomaly detection framework for residential buildings composed of a one-step ahead load predictor and a quantile regress-based load anomaly detector is presented in this study. Our research contributions are as follows:

- Deep learning method RNN is adopted to forecast the time series building energy consumption, this method has been applied in many literature [26–29], and shown great advantages in forecasting accuracy. Moreover, the Hatemi-J criterion (HJC) is employed to reduce the influence of the over fitting problem of the time series model.
- Quantile regression is a promising choice for the interpretability of anomaly detection results, as it exposes relationships between different quantile predictors and the target variable [30], and it is easy to judge whether the anomaly is within the given range. To the best of the authors' knowledge, it is the first at-

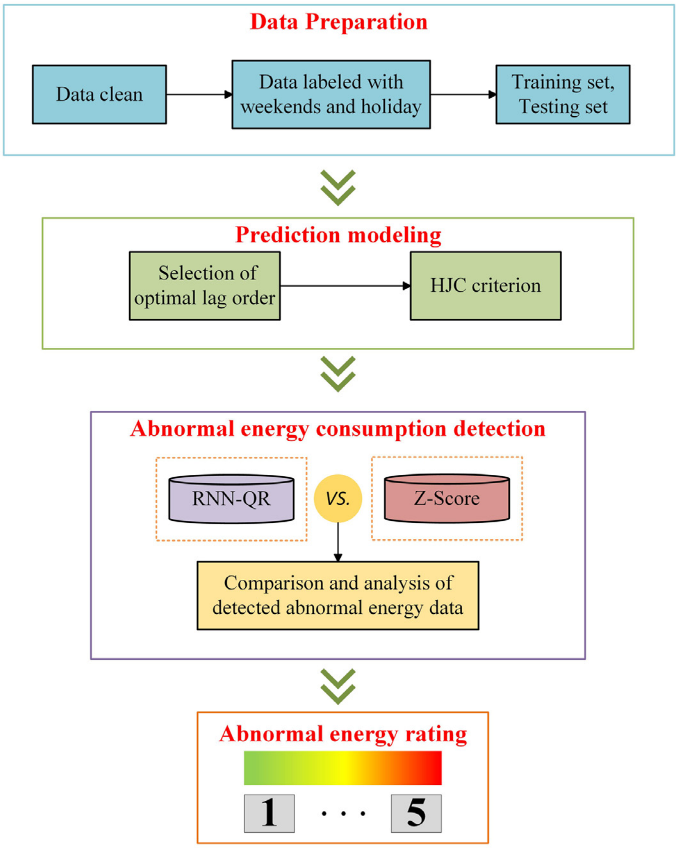

Fig. 1. Implementation flowchart for anomaly detection and evaluation.

tempt to combine RNN and quantile regression for anomaly detection.
- Three different residential houses are tested by using the proposed anomaly detection framework, and the detected results are divided into different grades based on different quantiles. The rationalization of the detected results are also discussed in depth from the temporal context.

The structure of this paper is as follows: Section 2 introduces the research framework of this paper, including data preparation, HJC criterion, RNN combined with QR method, and evaluation metrics used in this study. In Section 3, we summarize the results of anomaly detection and then discuss the rationality of them, moreover, evaluation and grading methods is discussed in detail. Conclusions are rendered at the end of the paper.

2. Methodology

This section gives a detailed description of the approach to detect the abnormal energy consumption in residential houses. As the traditional anomaly detection methods always evaluate a predefined threshold to detect anomalies, which is not suitable for complicated residential electricity consumption. Hence, the recurrent neural network with quantile regression (RNN-QR) model is utilized to detect the anomalies in residential houses in this paper, and to the best of the author's knowledge this type of model has not been applied in this area so far.

The flowchart of the proposed RNN-QR anomaly detection method is shown in Fig. 1. In this section, we provide more details on the implementation of the whole anomaly detection and evaluation process. There are three main parts that are described and discussed in this section: data preparation, HJC criterion, and RNN-QR.

2.1. Data preparation

The dataset is collected from different residential houses located in Burnaby in British Columbia, Canada [31], and it is obtained from an open source data platform that any customer can logged into the web and requested an export of historical hourly consumption data. In this work, we select dataset collected over a period of two years, ranging from January 2016 to January 2018. In addition, we have used three different houses from this dataset with house ids 3, 4, 5, and defined them as residential house1, residential house2, and residential house3 respectively.

In general, popular methods to deal with missing data mainly include moving window and inference-based methods [32]. Moving window method is essentially an improved model of linear interpolation, which is easy to implement when the duration of missing values is short. While for long duration of missing values, inference-based methods are recommended for analysis. In this study, to handle missing data in the raw dataset, we do it from two aspects: (1) for cases where the number of missing data is less than 3, moving window method is adopted as the changes in energy consumption are assumed to be linear within a short time; (2) for cases where the number of missing data is more than 3, we replace the missing data with the data at the same time the day before. Furthermore, 80% samples of datasets are used for training, and the remaining samples (20%) are reserved for testing.

2.2. HJC criterion

In this work, we use historical energy consumption to predict future energy consumption, thus it is vital to choose the optimal lag order of the time series model. In the literature a number of information criteria are available for finding the optimal lag order of time series model, and obtaining the best balance between the complexity and the accuracy of time series model simultaneously [33]. Two of the most widely used criteria presented in the literature are Bayesian information criterion (BIC) [34] and the Hannan and Quinn information criterion (HQC) [35]. These two criteria can be calculated as follows:

$$BIC = k \cdot \ln(n) - 2\ln(L) \tag{1}$$

$$HQC = k \cdot \ln(\ln(n)) - 2\ln(L) \tag{2}$$

Here, k is the lag order used in time series model, n is the sample size, and L represents the maximum likelihood estimate of the model.

However, there are situations when these two information criteria pick two different optimal lag orders, thus it is difficult to choose a versatile criterion. In order to solve this problem, Hatemi-J [33] combined these two criteria to obtain the following information criterion:

$$HJC = k \cdot (\ln(n) + \ln(\ln(n))) - 2\ln(L) \tag{3}$$

Moreover, simulation experiments show that the proposed information criterion can pick the true lag order in most cases. Therefore, HJC information criterion is applied in this work to obtain the optimal lag orders of time series models, and the results are shown in Section 3.1.

2.3. RNN-QR

A recurrent neural network is a class of artificial neural networks where connections between nodes form a directed graph

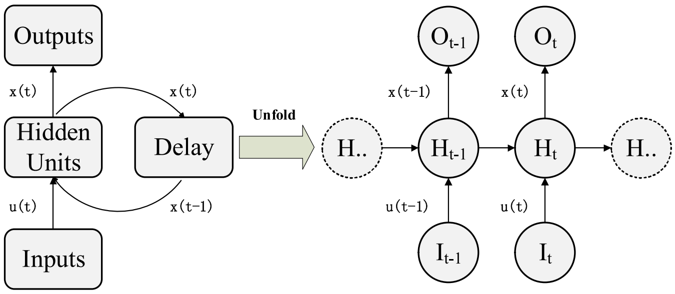

Fig. 2. Schematic of recurrent neural network.

along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. The memory captures information about what has been calculated so far [36]. A schematic of recurrent neural network is shown in Fig. 2. The mathematical presentation of RNN with delay of 1 is as follows:

$$x(t) = f_H(W_{IH}u(t) + W_{HH}x(t-1)) \tag{4}$$

$$O(t) = f_o(W_{HO}x(t)) \tag{5}$$

where the neural network inputs and outputs are the vectors u(t) and O(t), the three connection weight matrices are W_{IH}, W_{HH} and W_{HO}, and the hidden and output unit activation functions are f_H and f_o, respectively.

Based on the RNN prediction results, quantile regression is applied to obtain the probabilistic forecasting results. Quantile regression is a modeling method that models the conditional distribution quantile of the response variable as a function of the observed covariates [30]. Traditional regression models focus on finding the conditional mean of the target variable by minimizing the sum of squared residuals, while quantile regression finds the conditional median of the target variable by finding the sum of absolute residuals. Quantile regression can provide more information about future uncertainties by utilizing quantile loss function, which is also known as the pinball loss, and it is calculated as follows:

$$L_{q,t}\left(y_t, \hat{y}_t^q\right) = \sum_{i:y_t \le \hat{y}_t^q}^{N} (1-q)\left(\hat{y}_t^q - y_t\right) + \sum_{i:y_t > \hat{y}_t^q}^{N} q\left(y_t - \hat{y}_t^q\right) \tag{6}$$

where q denotes the targeted quantile, y_t is the target variable at time t, $\hat{y}_t^q$ denotes the estimated q_{th} quantile at time t, and $L_{q,\,t}$ denotes the pinball loss for the q_{th} quantile at time t.

To sum up, quantile regression can comprehensively describe the whole conditional distribution of explained variables, and reduce the influence of outliers on the predicted results. Furthermore, quantile regression can obtain a series of quantiles to represent the uncertainties and guarantee the rationality of the range of prediction results. It is therefore feasible to use quantile regression for abnormal energy consumption detection.

2.4. Evaluation metrics

In order to assess the results of the forecasting models based on quantile regression, two evaluation metrics obtained from [37] can be used in this work. Firstly, the PICP evaluates whether the actual value is within the calculated prediction interval limits. However, it become clear that it's not that the higher value of the PICP, the better, as a very wide prediction interval may lead to meaningless results. Therefore, a narrow prediction interval as well as high PICP should be the main purpose of the forecast. The second evaluation

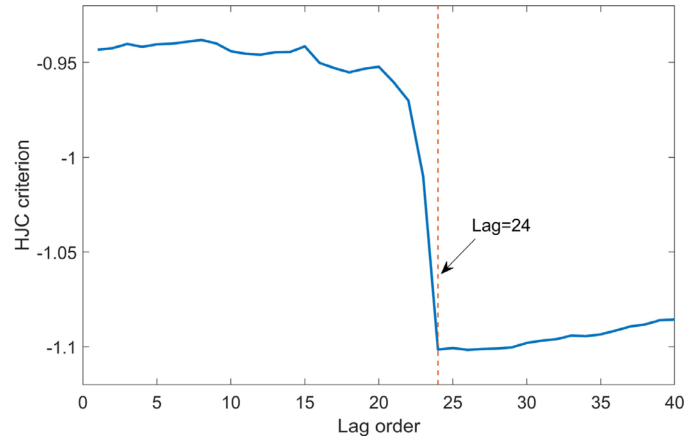

Fig. 3. Optimal lag order curve for maximum 40 lags in residential house 1.

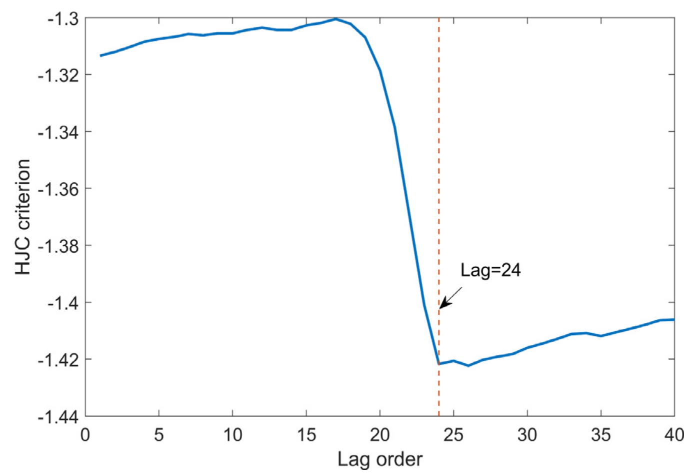

Fig. 4. Optimal lag order curve for maximum 40 lags in residential house 2.

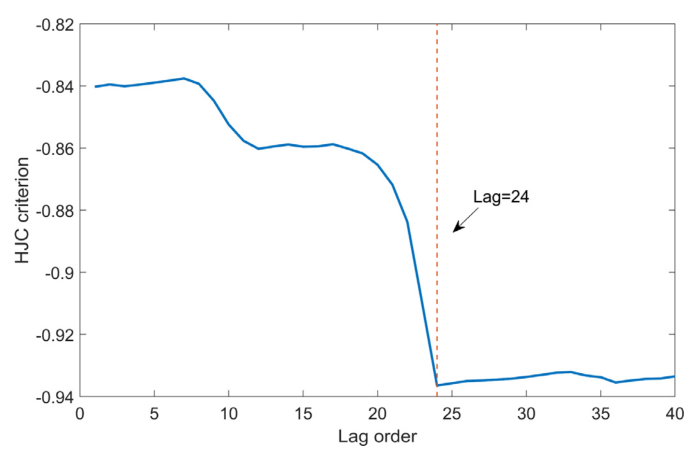

Fig. 5. Optimal lag order curve for maximum 40 lags in residential house 3.

metric PINAW can measure the width of the prediction interval. These two metrics are defined as below:

$$PICP = \frac{1}{N}\sum_{i=1}^{N}\alpha_i \tag{7}$$

$$PINAW = \frac{1}{NE}\sum_{i=1}^{N}(U_i - L_i) \tag{8}$$

where $\alpha_i = 1$ if the actual value lies within the prediction interval, and $\alpha_i = 0$ otherwise. L_i and U_i represent the lower and the upper

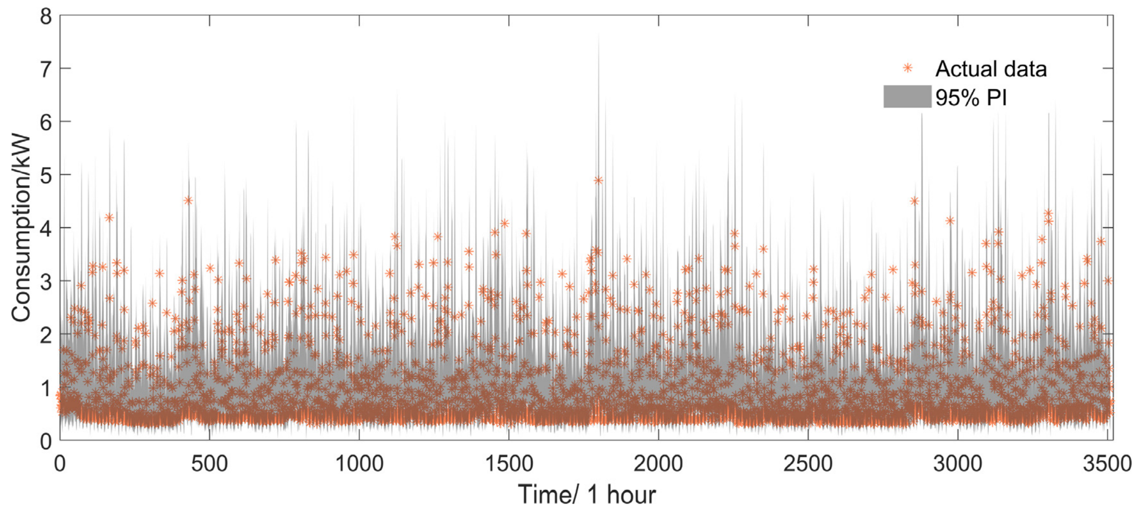

Fig. 6. The coverage of the 95% quantile interval over the testing data from residential house 1.

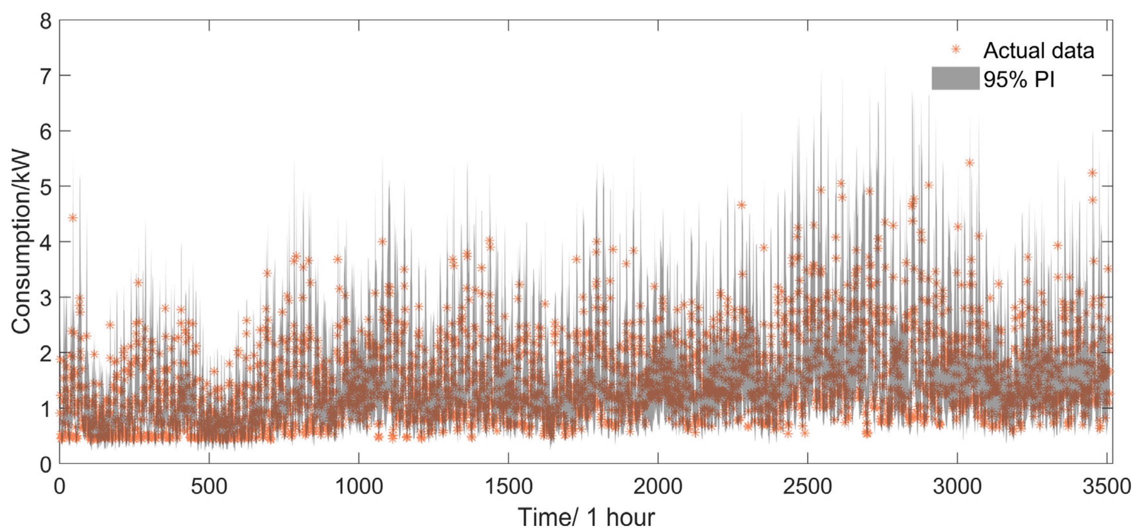

Fig. 7. The coverage of the 95% quantile interval over the testing data from residential house 2.

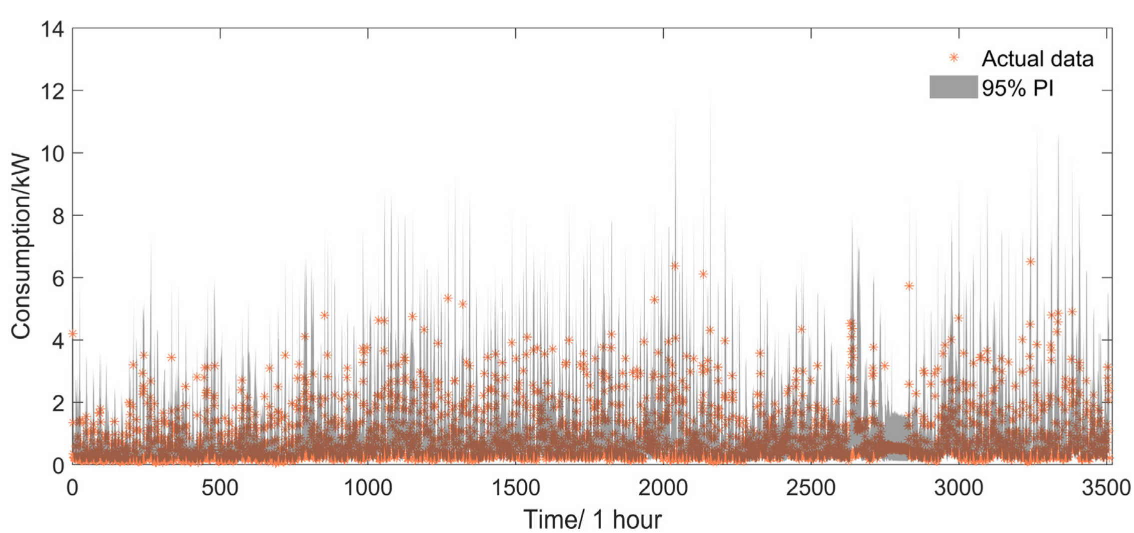

Fig. 8. The coverage of the 95% quantile interval over the testing data from residential house 3.

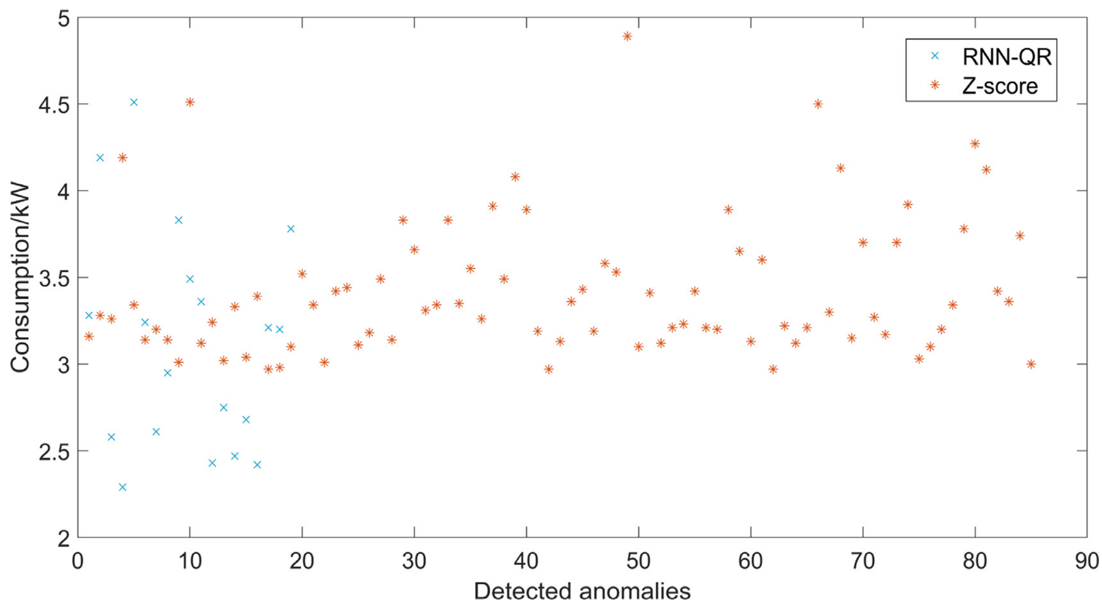

Fig. 9. The distribution of anomalies detected by RNN-QR and Z-score in residential house 1.

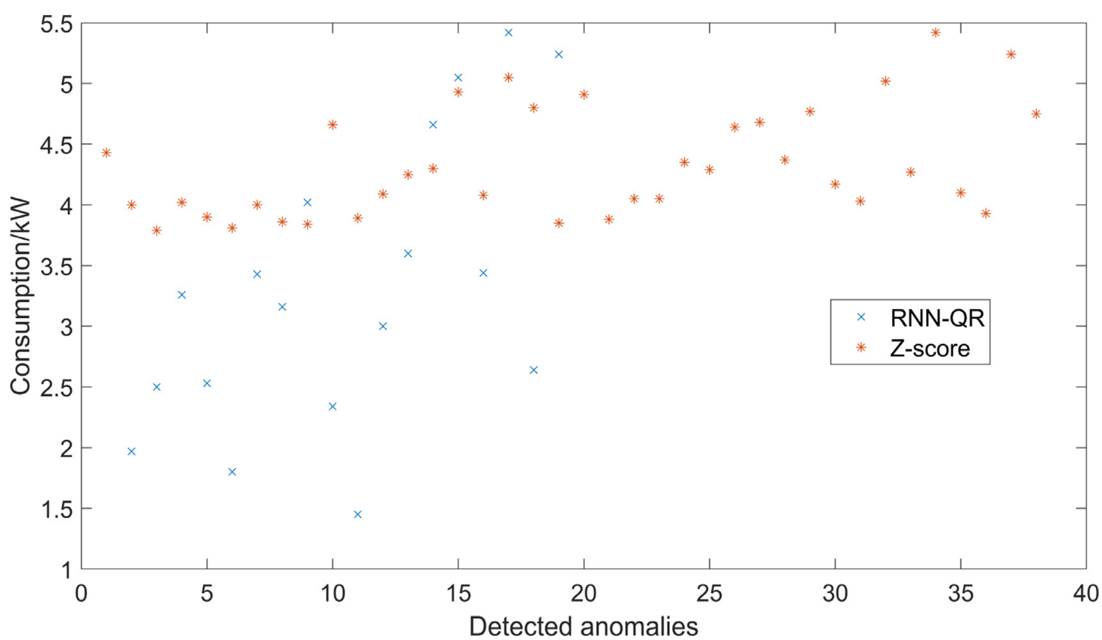

Fig. 10. The distribution of anomalies detected by RNN-QR and Z-score in residential house 2.

boundaries of the prediction interval respectively. E is the difference between the maximum and the minimum actual values, and its main role is to normalize the result.

3. Results and discussion

3.1. Anomaly detection results

In this work, anomaly detection experiments on real-world public datasets from three residential houses mentioned above are conducted. Firstly, HJC criterion is adopted to obtain the optimal lag order of the prediction model, and the results are illustrated from Fig. 3 to Fig. 5. It can be noticed that residential house 1 to 3 have the same optimal lag order 24, and this might be explained by the fact that residents' electricity use habits at this time are similar to those at the same time the previous day. Whereafter, RNN prediction models with 10-fold cross validation are established based on the optimal lag order. Particularly, in order to verify the generalization of the proposed models, the structure of these models for different residential houses are set to be identical. Finally, quantile regression is applied to build probabilistic power consumption forecasting models, we select quantile interval of five different ranges, e.g. 95%, 96%, 97%, 98%, 99%, and the rarity of events increases as the range of quantile interval increases, therefore we can define these rare events as detected anomalies.

The coverage of the 95% quantile interval over the testing data from the three residential houses are exhibited in Fig. 6 to Fig. 8. As we can see from the figures, almost all the actual electricity consumption is well captured by the proposed models. Furthermore, different residential houses follow different energy usage habits, and for most of the time, the electricity consumption of residential houses is no more than 2 kW, while the peak demand for

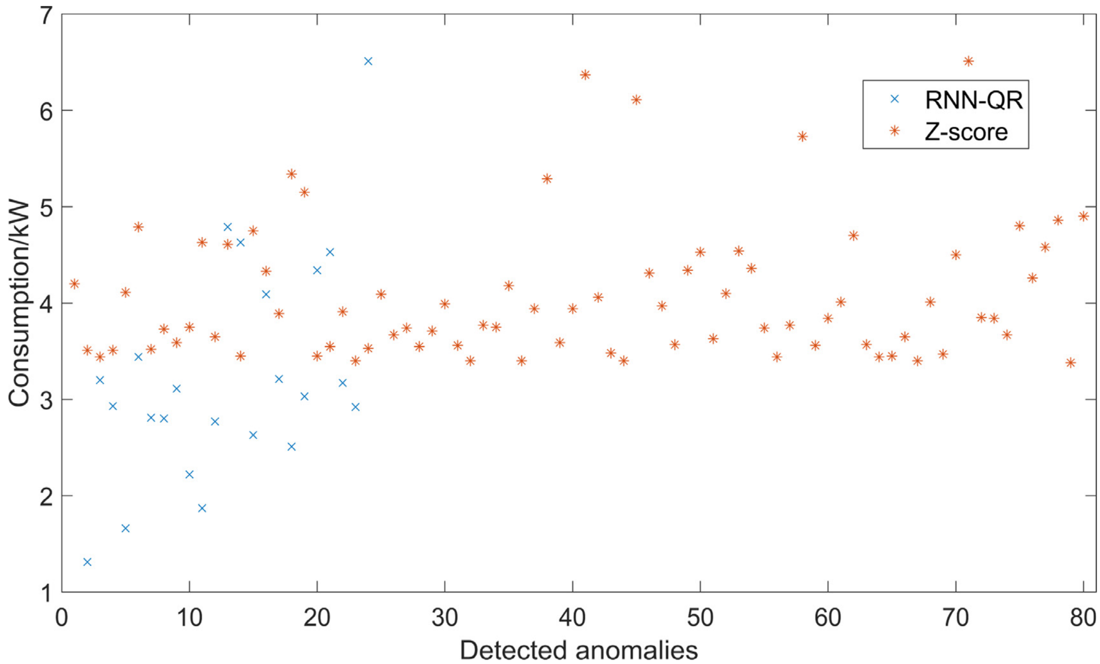

Fig. 11. The distribution of anomalies detected by RNN-QR and Z-score in residential house 3.

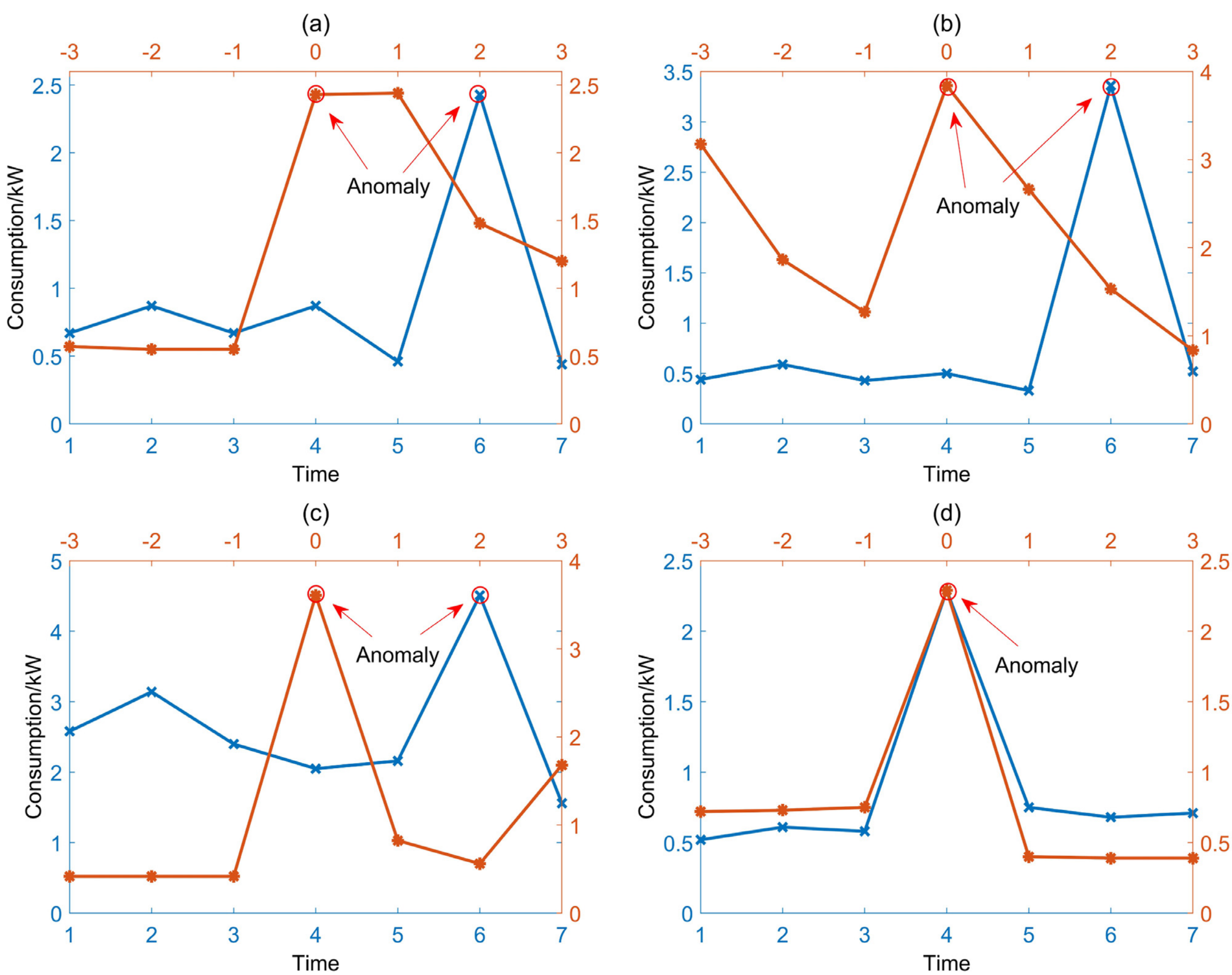

Fig. 12. Anomaly detection results of residential house 1, (a) 6 a.m. (b) 12 a.m. (c) 4 p.m. (d) 9 p.m.

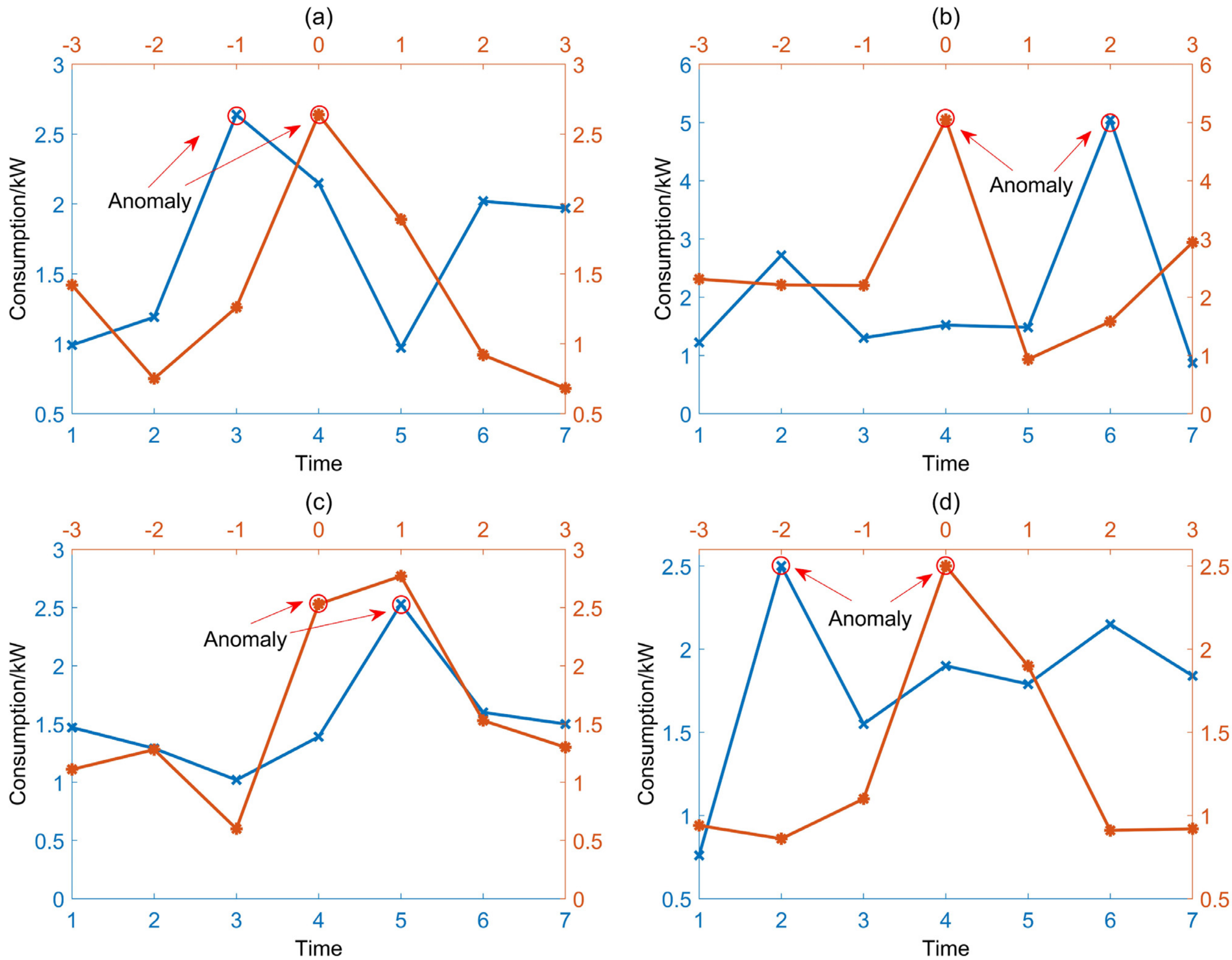

Fig. 13. Anomaly detection results of residential house 2, (a) 10 a.m. (b) 1 p.m. (c) 4 p.m. (d) 7 p.m.

electricity consumption is between 6 kW and 8 kW. Moreover, it is worth noting that the time span of the above figures are set to be the same, which makes the prediction results more comparable.

Table 1 shows the numeric results of electricity consumption prediction accuracy for residential houses. It is obviously that the variation range of electricity consumption is accurately predicted by RNN-QR models, and almost 97% of the actual data falls into the provided 95% quantile interval. In addition, the value of PINAW are achieved between 36.1% and 43.9%, which can avoid the situation that the uncertainty information of the prediction value cannot be given effectively by excessively wide quantile interval.

Table 1
Electricity prediction accuracy for residential houses.

Model	PICP	PINAW
Residential house1	0.974	0.439
Residential house2	0.972	0.377
Residential house3	0.976	0.361

Table 2
Number of detected anomalies under 99% quantile interval.

Model	RNN-QR	Z-score
Residential house1	19	85
Residential house2	19	38
Residential house3	24	80

3.2. Comparison with Z-score

The proposed models compared with Z-score models is employed in this research. For the purpose of simplifying comparisons, we select testing data falling outside the 99% quantile interval as the detected anomalies. For quantile regression, the 99% quantile interval is consist of 0.5% and 99.5% quantile, while for Z-score models, the 99% quantile interval lies within three standard deviations above and below the mean value of datasets.

Table 2 summarizes the number of detected anomalies under 99% quantile interval. Obviously, the anomalies detected by Z-score method are much more than those detected by RNN-QR method. It is partial to judge the model only from the number of anomalies, therefore, the distribution of detected anomalies are depicted in Fig. 9 to Fig. 11. On the one hand, it can be seen that almost all of the anomalies detected by Z-score fall above the line $y = 3$, which means that only high power consumption can be identified as anomalies. Thus, the main drawback of Z-score method is that it only focuses on high power consumption and ignores the possible unreasonable low power consumption. The anomalies detected by RNN-QR method, on the other hand, form more reasonable distributions as presented from Fig. 9 to Fig. 11. It is clear that there is no fixed upper bound or lower bound in RNN-QR method, and the quantile interval of RNN-QR changes along with

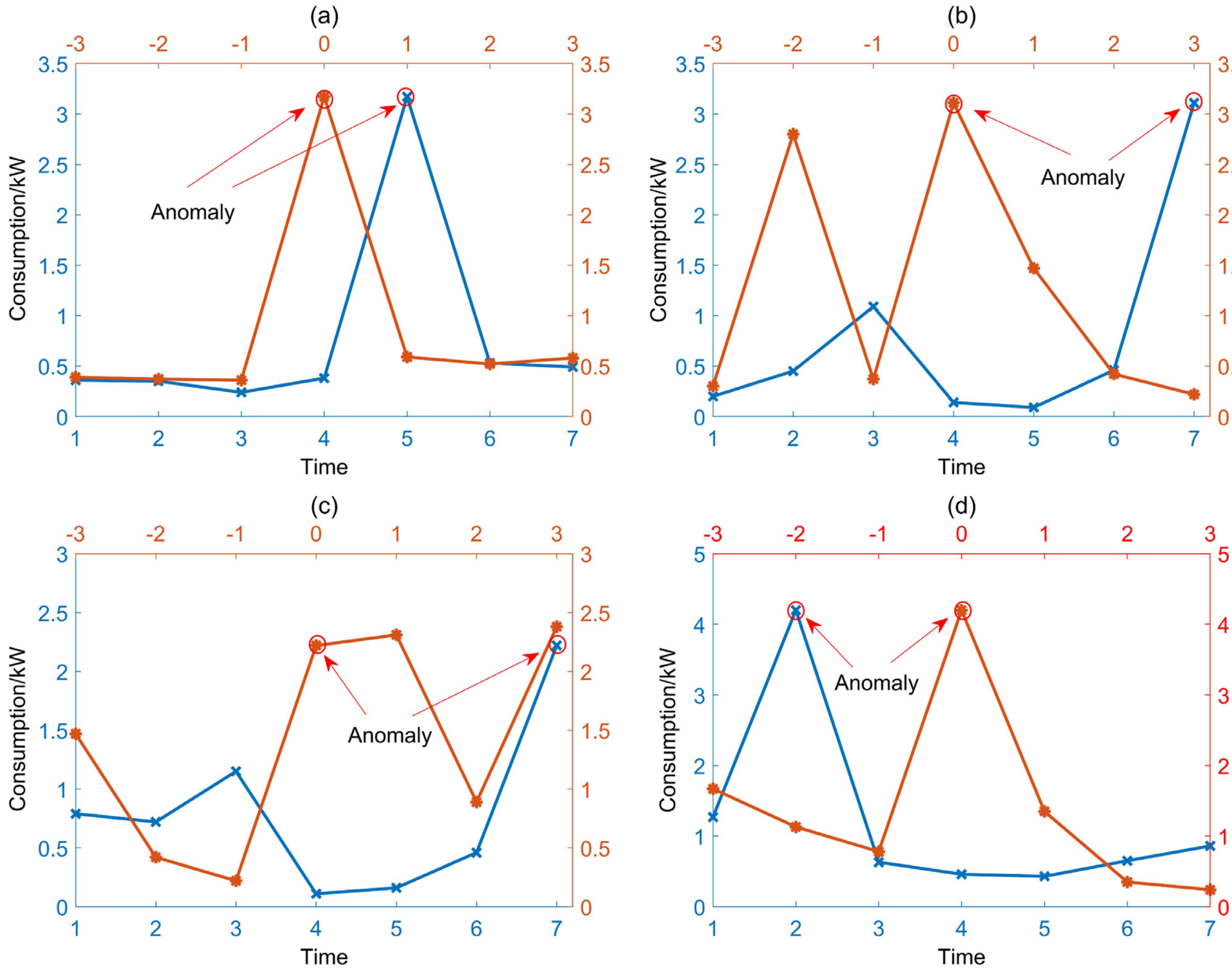

Fig. 14. Anomaly detection results of residential house 3, (a) 7 a.m. (b) 11 a.m. (c) 3 p.m. (d) 8 p.m.

the variety of actual value, which makes the detected results more interpretable.

To demonstrate the advantages of RNN-QR method and verify the rationality of detected anomalies, we provide a detailed comparison from two aspects: (1) compared the detected anomalies with the actual data within several hours before and after them; (2) compared the detected anomalies with the detected anomalies. The first aspect focus on changes in power consumption over a continuous period of time, as in most definitions, anomalies are values that deviate greatly from other data in a certain time period. For the second aspect, we assume that energy usage habit of residents at the same time within a week will not change dramatically.

Fig. 12 to Fig. 14 present randomly selected detected results from three residential houses, and in these figures, red lines represent the hourly continuous energy consumption values, while blue lines represent energy consumption values at the same time within a week. In general, the rationality of detected anomalies is sufficient as we can infer from the above two aspects. Furthermore, the proposed approach may have a potential advantage, i.e., finding anomalies in frequent fluctuations of power consumption. As illustrated in Fig. 12b, Fig. 14b and c, the actual data within several hours before and after the anomaly fluctuates frequently, it is therefore difficult to judge whether there is any anomaly from this point of view. However, when the time span expands to one week, it is clear that anomaly occurs exactly when energy usage habit of residents change the most in a week.

However, it must be realized that simply comparing the energy consumption with previous patterns does not fully explain the rationality of the anomalies, for example, the occasional activities like family dinner may cause high energy consumption. This defect is not fatal in practical applications. On the one hand, residents can realize the increase of energy consumption when using household appliances occasionally, hence the alerts given by the anomaly detection model will not make them feel puzzled. On the other hand, only about 20 anomalies are detected in the testing set, and these are relatively rare event for a long time span of nearly five months. In general, the main aim of the proposed methodology is to remind residents of possible energy consumption anomalies, so as to reduce energy waste.

As stated above, compared with the Z-score method, the proposed method shows obvious advantages in detecting abnormal energy consumption. For anomalies that fall outside the range of 99% quantile interval, it is worth noting that different from Z-score method, which only focuses on high energy consumption events, the proposed method is effective in the detection of relatively low energy consumption events, thus the distribution of anomalies detected by the proposed method is reasonable. In addition, through the above discussion on the rationality of anomalies, it can be inferred that the detected anomalies is reasonable and interpretable.

3.3. Abnormal energy grading

In this section, we conduct a comprehensive analysis of grading the detected anomalies. It is easy to think of matching quantiles

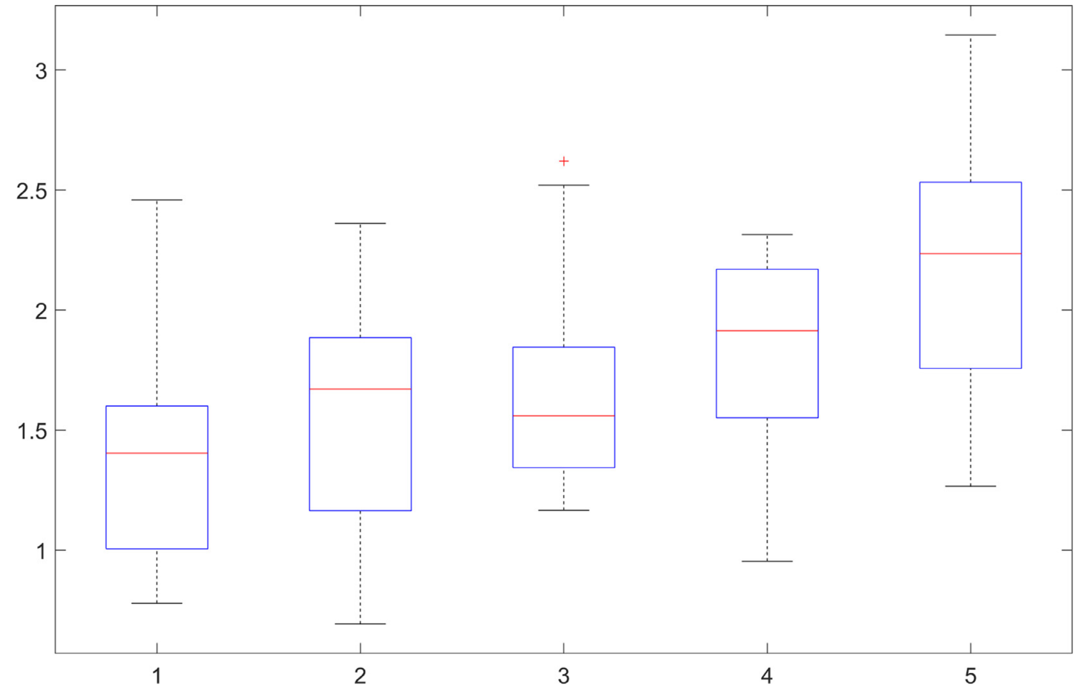

Fig. 15. Distribution of difference between anomaly and predicted value in residential house 1.

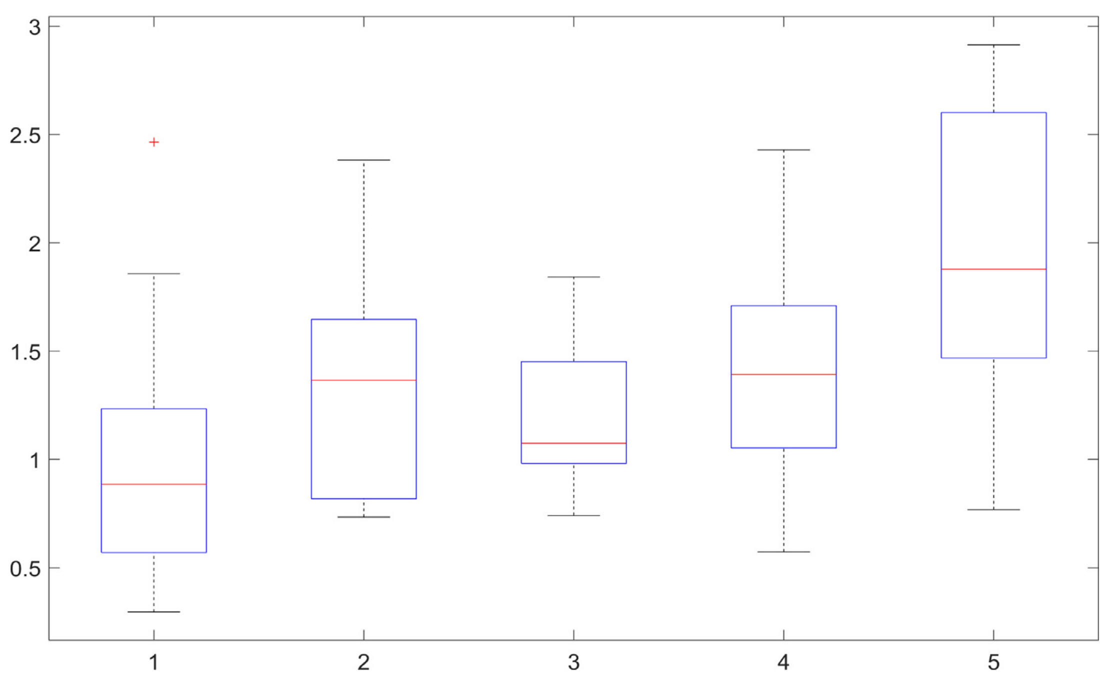

Fig. 16. Distribution of difference between anomaly and predicted value in residential house 2.

with grades. Therefore, as mentioned in Section 3.1, we select quantile interval of five different ranges from 95% to 99%, and define level 1 as anomalies between quantile 95% and 96%, level 2 as anomalies between quantile 96% and 97%, and so on. Specially, level 5 stand for the anomalies fall outside the 99% quantile interval.

In order to improve visibility of rating results and verify their rationality, we calculate the difference between each anomaly and its predicted value in each grade, and the results are presented in the form of box-plot in Fig. 15 to Fig. 17. It can be seen that, in general, the higher grade anomalies correspond to the larger deviation between actual value and predicted value. However, it is worth noting that the width of the quantile interval is not constant, thus higher grade anomalies are not necessarily larger than lower grade anomalies in terms of deviations between anomalies and predicted value, and these three figures only reveal the trend of deviations of anomalies from predicted values. More specially, the anomalies of grade 2 do not conform to the aforementioned rule, as depicted from Fig. 15 to Fig. 17, the deviations of grade 2 is generally greater than grade 3. A possible reason for this is that the width of the quantile interval of anomalies in grade 2 is relatively large, which leads to a relatively low grade for the same degree of deviation.

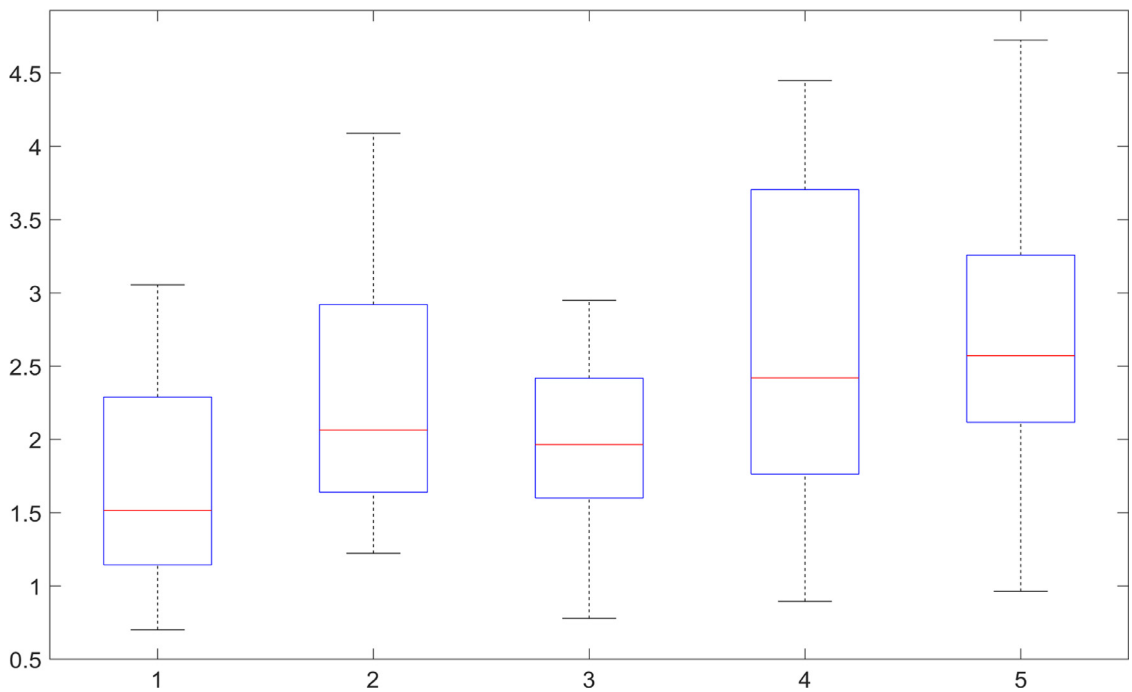

Fig. 17. Distribution of difference between anomaly and predicted value in residential house 3.

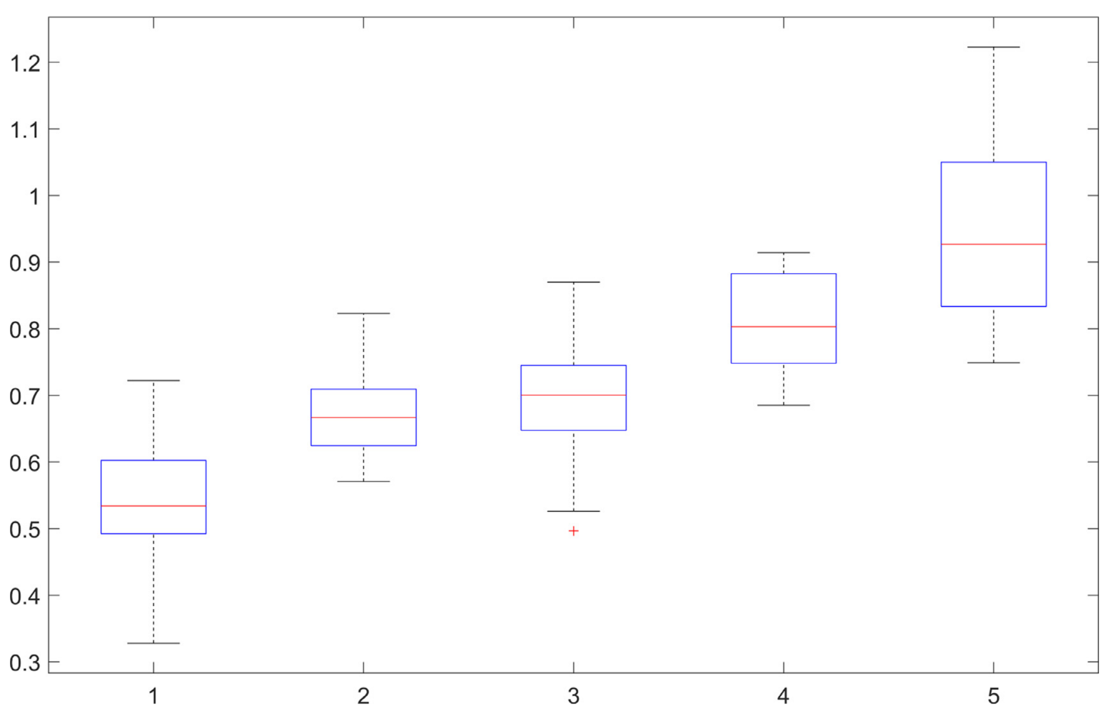

Fig. 18. Distribution of the proposed ratio in residential house 1.

In order to mitigate the influence of the width of the quantile interval, we provide a more reasonable comparison for different grades. As mentioned above, the width of the quantile interval is not constant for different anomalies, we therefore calculate the ratio of the difference between the anomaly and its median prediction result to its width of 99% quantile interval, and it is a more convincing strategy to use the ratio as the comparison criterion. Fig. 18 to Fig. 20 show the distributions of ratio in five grades for three different residential houses. Similar to the above, a visual inspection of these figures show that, on the whole, the higher grade anomalies correspond to the larger ratio. It can be concluded that the proposed comparison criterion can fully reflect the difference of different grades for detected anomalies. Therefore, the rationality of grading the detected anomalies can be verified from two aspects: (1) for a single detected anomaly, grading is based on the results of quantile regression, which has been discussed at the beginning of this subsection; (2) for all detected anomalies, we propose an appropriate comparison criterion, which reveals the severity of the anomalies to some extent. Based on the proposed criterion, the distribution of anomalies is basically consistent with the results of grading, and this indicates that the grading method is applicable to all detected anomalies.

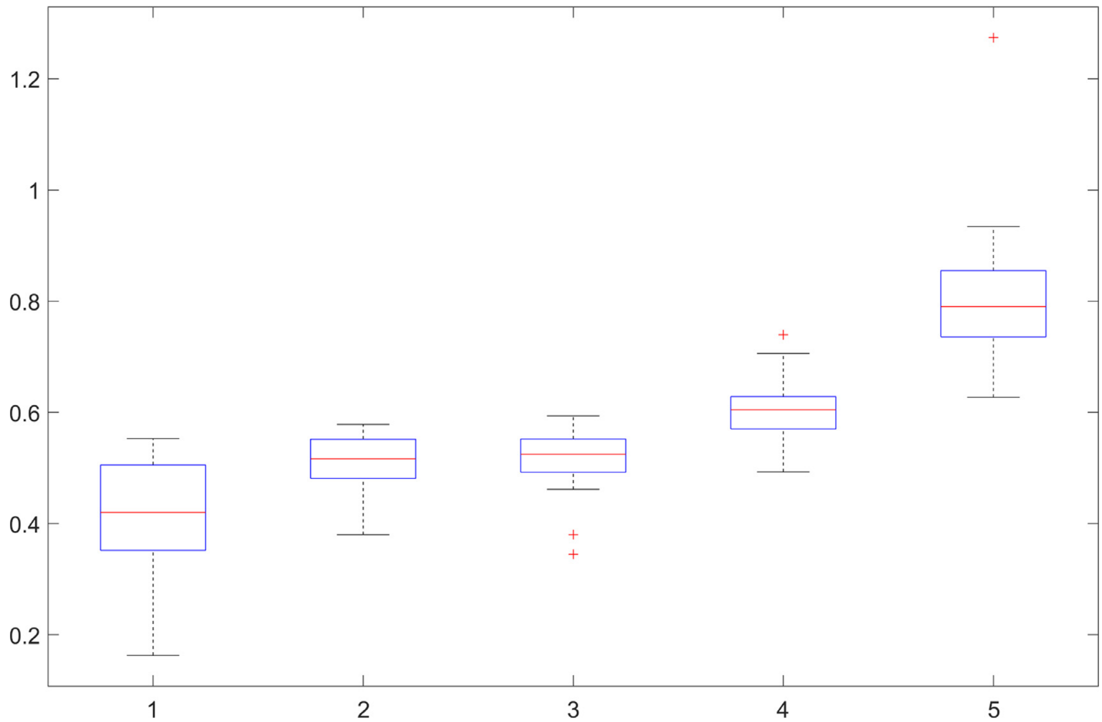

Fig. 19. Distribution of the proposed ratio in residential house 2.

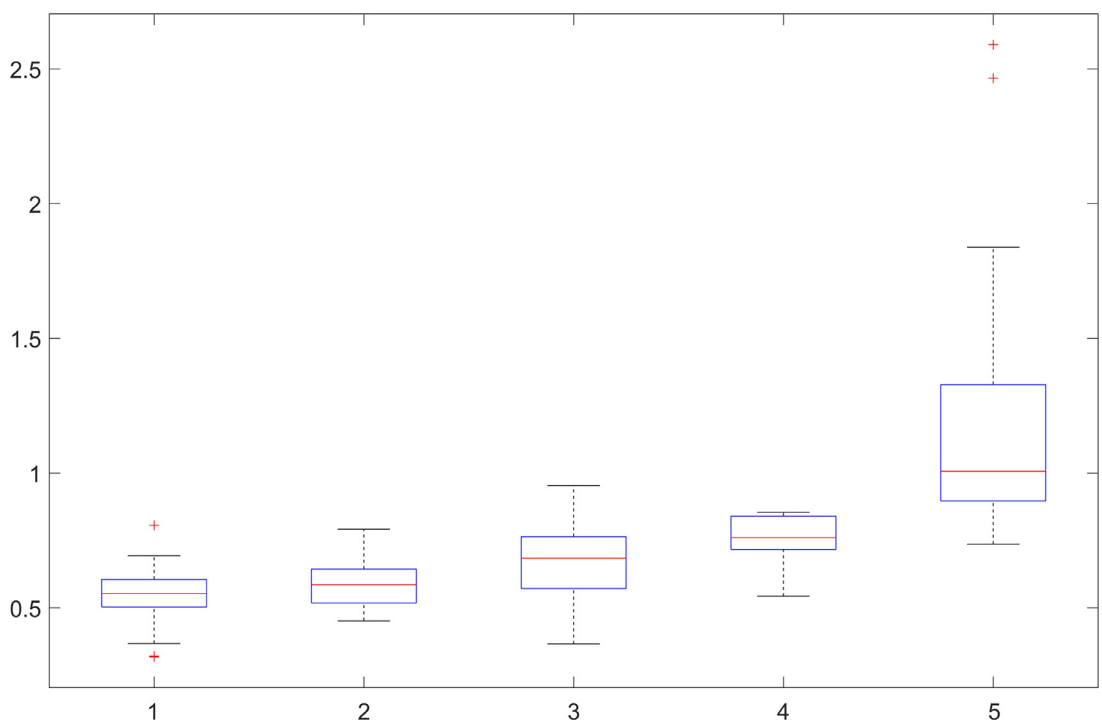

Fig. 20. Distribution of the proposed ratio in residential house 3.

4. Conclusion

Anomaly detection can help the building operation managers find abnormal energy consumption behavior and equipment failure during the system operation, so as to improve the performance of the building and reduce unnecessary energy waste. However, for most of the practical applications, it is impossible to collect metadata with anomaly labels, so how to accurately detect the anomaly with limited information is the difficulty of the research. The main contribution of this work is summarized as follows: (1) it is the first attempt to combine the deep learning method RNN with quantile regression for anomaly detection, and the rationality of the detected anomaly has been verified; (2) the detected anomalies are graded according to the results of the quantile regression, which could be useful in assessing the severity of anomalies.

To illustrate the advantages of the proposed model, it is compared with the widely used Z-score method. The results show that the anomalies detected by the proposed method are more practicable than that detected by Z-score method, as the proposed method does not simply divide the energy consumption that higher than a certain threshold into anomalies. In addition, the rationality of the anomalies detected by the proposed method is validated from two time spans: several hours and one week. The former represents the occasional energy-intensive behavior of residents, such as the sudden use of multiple high-power electrical appliances, while the latter represents changes in residents' energy usage habits, which may include equipment failures or changes to family plans. It is concluded that the detected anomalies are reasonable and interpretable.

As for the grading of anomalies, we classify anomalies into five different grades according to the quantile range they fall into. The anomaly grades reflect the ratio of the difference between the anomaly and its median prediction result to its width of 99% quantile interval, thus this grading method can reasonably be used to assess the severity of the anomaly. Furthermore, it is worth noting that the proposed anomaly detection model has been applied and verified in three different residential houses, which indicates that this model has strong ability of robustness and generalization. Further work will take into account the impact of weather and occupancy on abnormal energy consumption in residential houses.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

CRediT authorship contribution statement

Chengliang Xu: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Huanxin Chen:** Supervision.

Acknowledgement

The authors gratefully acknowledge the support of National Natural Science Foundation of China (Grant Nos. 51876070 and 51576074).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.enbuild.2020.109864.

References

[1] M. Ma, W. Cai, Y. Wu, China act on the energy efficiency of civil buildings (2008): a decade review, sci total environ, 651 (Pt 1) (2019) 42–60.
[2] J.Y. Park, Z. Nagy, Comprehensive analysis of the relationship between thermal comfort and building control research - A data-driven literature review, Renew. Sustain. Energ. Rev. 82 (2018) 2664–2679.
[3] H.L. Wang, P. Xu, X. Lu, D.K. Yuan, Methodology of comprehensive building energy performance diagnosis for large commercial buildings at multiple levels, Appl Energ 169 (2016) 14–27.
[4] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Syst. Appl. 42 (9) (2015) 4324–4338.
[5] D.B. Araya, K. Grolinger, H.F. ElYamany, M.A.M. Capretz, G. Bitsuamlak, An ensemble learning framework for anomaly detection in building energy consumption, Energ. Buildings 144 (2017) 191–206.
[6] Y.B. Guo, Z.H. Tan, H.X. Chen, G.N. Li, J.Y. Wang, R.G. Huang, J.Y. Liu, T. Ahmad, Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving, Appl. Energ. 225 (2018) 732–745.
[7] G.N. Li, H.X. Chen, Y.P. Hu, J.Y. Wang, Y.B. Guo, J.Y. Liu, H.R. Li, R.G. Huang, H. Lv, J. Li, An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators, Appl. Therm. Eng. 129 (2018) 1292–1303.
[8] J.Y. Liu, J.H. Liu, H.X. Chen, Y. Yuan, Z.F. Li, R.G. Huang, Energy diagnosis of variable refrigerant flow (VRF) systems: data mining technique and statistical quality control approach, Energ. Buildings 175 (2018) 148–162.
[9] J.S. Chou, A.S. Telaga, Real-time detection of anomalous power consumption, Renew. Sust. Energ. Rev. 33 (2014) 400–411.
[10] B. Rosner, Percentage points for a generalized ESD many-outlier procedure, Technometrics 25 (2) (1983) 165–172.
[11] J.E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energ. Buildings 39 (1) (2007) 52–58.
[12] J.E. Seem, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, Energ. Buildings 37 (2) (2005) 127–139.
[13] C. Fan, F. Xiao, S.W. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Appl. Energ. 127 (2014) 1–10.
[14] U. Habib, G. Zucker, M. Blochle, F. Judex, J. Haase, Outliers Detection Method Using Clustering in Buildings Data, Iecon 2015 - 41st Annual Conference Of the Ieee Industrial Electronics Society, (2015) 694–700.
[15] N. Cotrufo, R. Zmeureanu, PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers, Energ. Buildings 130 (2016) 443–452.
[16] M. Molina-Solana, M. Ros, M.D. Ruiz, J. Gomez-Romero, M.J. Martin-Bautista, Data science for building energy management: a review, Renew. Sust. Energ. Rev. 70 (2017) 598–609.
[17] Y. Zhao, T. Li, X. Zhang, C. Zhang, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future, Renew. Sustain. Energ. Rev. 109 (2019) 85–101.
[18] Y. Sun, T.Y. Wu, G.T. Zhao, M. Guizani, Efficient rule engine for smart building systems, IEEE T Comput. 64 (6) (2015) 1658–1669.
[19] Z. Yu, F. Haghighat, B.C.M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data, Energ. Buildings 47 (2012) 430–440.
[20] D.F.M. Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes, Energ. Buildings 62 (2013) 210–216.
[21] C. Fan, F. Xiao, Y. Zhao, J.Y. Wang, Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data, Appl. Energ. 211 (2018) 1123–1135.
[22] G. Mavromatidis, S. Acha, N. Shah, Diagnostic tools of energy performance for supermarkets using artificial neural network algorithms, Energ. Buildings 62 (2013) 304–314.
[23] L. Yan, P.F. Hu, C.H. Li, Y. Yao, L. Xing, F. Lei, N. Zhu, The performance prediction of ground source heat pump system based on monitoring data and data mining technology, Energ. Buildings 127 (2016) 1085–1095.
[24] Y. Zhao, S. Wang, F. Xiao, A statistical fault detection and diagnosis method for centrifugal chillers based on exponentially-weighted moving average control charts and support vector regression, Appl. Therm. Eng. 51 (1–2) (2013) 560–572.
[25] X. Liu, P.S. Nielsen, Scalable prediction-based online anomaly detection for smart meter data, Inf. Syst. 77 (2018) 34–47.
[26] D. Koschwitz, J. Frisch, C. van Treeck, Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX recurrent neural network: a comparative study on district scale, Energy 165 (2018) 134–142.
[27] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energ. 212 (2018) 372–385.
[28] Z. Afroz, T. Urmee, G.M. Shafiullah, G. Higgins, Real-time prediction model for indoor temperature in a commercial building, Appl. Energ. 231 (2018) 29–53.
[29] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energ. 236 (2019) 700–710.
[30] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge; New York, 2005.
[31] S. Makonin, HUE: the hourly usage of energy dataset for buildings in British Columbia, Data. Brief. 23 (2019) 103744.
[32] B.K. Vaughn, Data analysis using regression and multi-level/hierarchical models, J. Educ. Meas. 45 (1) (2008) 94–97.
[33] A. Hatemi-J, A new method to choose optimal lag order in stable and unstable VAR models, Appl. Econ. Lett. 10 (3) (2003) 135–137.
[34] A. Mehrjou, R. Hosseini, B.N. Araabi, Improved Bayesian information criterion for mixture model selection, Pattern Recogn. Lett. 69 (2016) 22–27.
[35] L. Gerencser, L. Finesso, A two-stage information criterion for stochastic systems revisited, Automatica 47 (12) (2011) 2791–2795.
[36] M. Paliwal, U.A Kumar, Neural networks and statistical techniques: a review of applications, Expert Syst. Appl. 36 (1) (2009) 2–17.
[37] M. Shepero, D. van der Meer, J. Munkhammar, J. Widén, Residential probabilistic load forecasting: a method using Gaussian process designed for electric load data, Appl. Energ. 218 (2018) 159–172.